



HELP

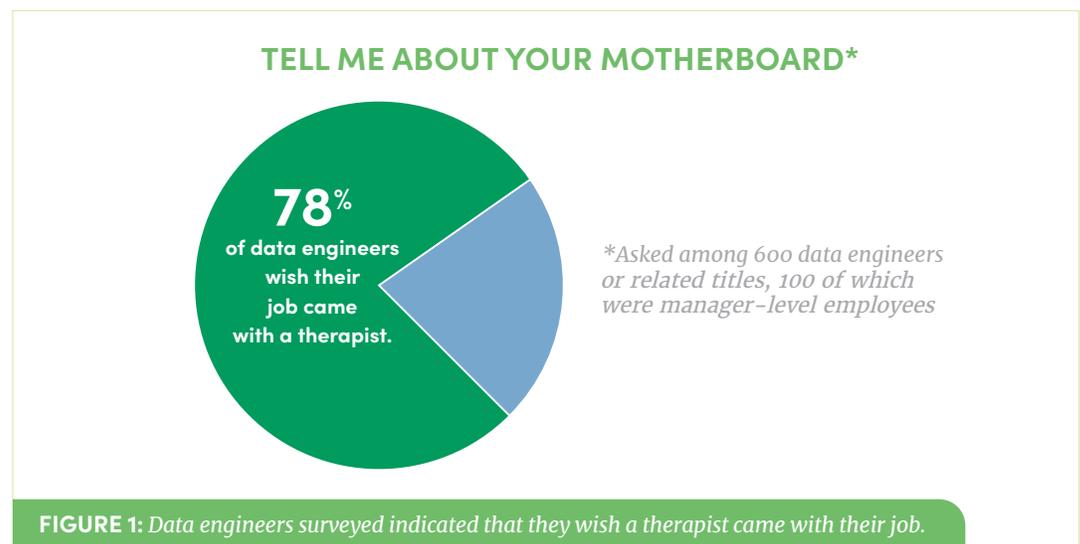
WHITE PAPER

2021 Data Engineering Survey

Burned-Out Data Engineers Call for DataOps

Data engineers are the backbone of the modern data-driven enterprise, and they serve in one of the most critical and celebrated roles in the tech industry. While data engineering tools and trends are frequently discussed and analyzed, conversations rarely focus on the day-to-day lives of data engineers and their lived experiences. To delve into this largely unexplored subject, [DataKitchen](#) and [data.world](#) teamed up to commission a survey intended to further understand the people behind the keyboards. Our survey gives voice to 600 data engineers and data engineering managers¹ who shared their hopes, challenges, frustrations and insight.

There are numerous challenges in the data analytics space. Recently, Gartner² wrote that “most analytics and AI projects fail because operationalization is only addressed as an afterthought.” The average tenure of a CDO or CAO is only about [2.5 years](#). Our survey confirms that data engineering is not immune to these systemic forces. A full 78% of those surveyed wished that their job came with a therapist to help manage work-related stress (see Figure 1). We share the insights from our survey in the hope that enterprises will heed the call to action and institute process, workflow and other supportive changes for data engineers and similar roles.



WHAT IS A DATA ENGINEER?

A data engineer is a software or computer engineer who lays the groundwork for team members, like analysts and data scientists, to perform analytics. Data engineers ensure that data is available, secure, correct, and fit for purpose. For example, the data engineer moves data from operational systems (ERP, CRM, MRP, etc.) or third-party sources into a data lake and writes the transforms that populate schemas in the data warehouses and data marts that power self-service analysis or automated charts, graphs and models.

Data engineering requires proficiency in cloud and cluster computing, ETL frameworks, batch/stream processing, orchestration, containers, coding, Agile Development, DevOps and observability. A competent data engineer influences the productivity of many others, so data engineers are valued and paid more on average than data analysts, scientists and DB admins. The life of a data engineer sounds pretty rosy when we look at it from thirty-thousand feet, but let's explore what data engineering is really like through the lens of our survey respondents.

THE DAY-TO-DAY WORK LIFE OF A DATA ENGINEER

When we asked data engineers how they feel about their daily work, 97% reported feeling *burned-out*. Data engineering is a relentlessly demanding profession in which you face a steady stream of requests from users, high-priority interruptions and ill-defined projects. The data engineers in our survey listed the challenges below as significant contributors to their feeling of burnout:

Sources of Burnout	Percent Responded
Focusing too much time on finding and fixing errors	50%
Focusing too much on maintaining data pipelines and/or manual processes	50%
Constantly playing catch up with stakeholder requests	49%
The fast pace of requests from stakeholders	48%
Lack of feedback on the products delivered	47%
Unreasonable requests from stakeholders	42%

TABLE 1: Sources of burnout identified by data engineers and managers

THE RELENTLESS FLOW OF ERRORS

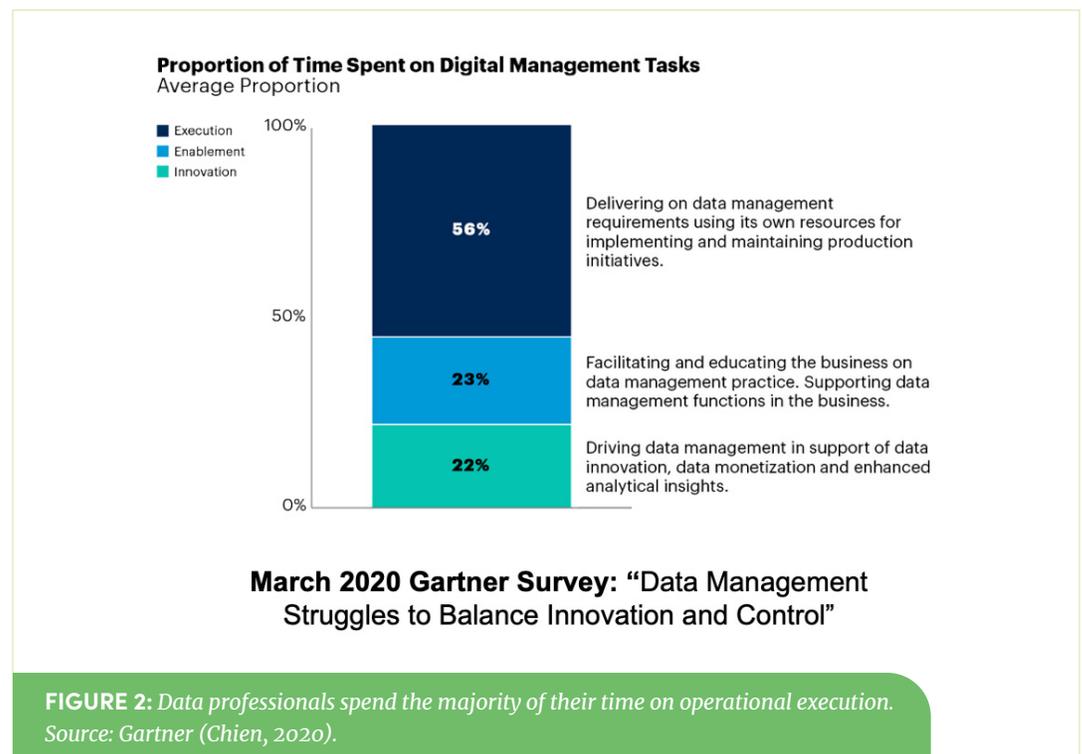
Data errors are a huge productivity drain on data engineers. Yet 52% of those surveyed did not feel like their companies sufficiently addressed data quality issues in a rigorous and systematic way. These respondents indicated that they “frequently hope and pray that things don’t break.” Many enterprises ingest data from sources and transform it with minimal quality controls. When working toward a deliverable, it is tempting to just quickly produce a solution with minimal testing, push it out to the users and hope it does not break. This approach has inherent risks. Eventually, a deliverable will contain data errors, upsetting the users and harming the hard-won credibility of the data analytics team.

A data engineer in a typical enterprise knows that errors can occur at any time. With no strategy to eliminate errors, the data engineering team can only “hope” for the best — another hour, another day — until the next interruption. A typical enterprise experiences multiple data, pipeline or analytics errors per week. Managing a continuing succession of outages while trying to keep development projects on schedule and under budget is like trying to play “*whack a mole*” while simultaneously reading a book. Data engineers know that an enterprise cannot derive value from its data unless the data team stays focused on innovation, but errors create unplanned work that can’t be ignored. For data engineers, this is a no-win situation.

MANUAL PROCESSES CROWD OUT INNOVATION

A recent Gartner³ survey showed that data professionals spent 56% of their time on operational execution and only 22% on innovation that delivers value — figure 2. Gartner describes the time spent on “operational execution” as using the data team to implement and maintain production initiatives. In other words, highly skilled data team members are asked to manually execute procedures that ingest, clean, transform, and disseminate data. Our survey confirms the magnitude of this problem, with 50% citing manual processes as an issue for data engineers.

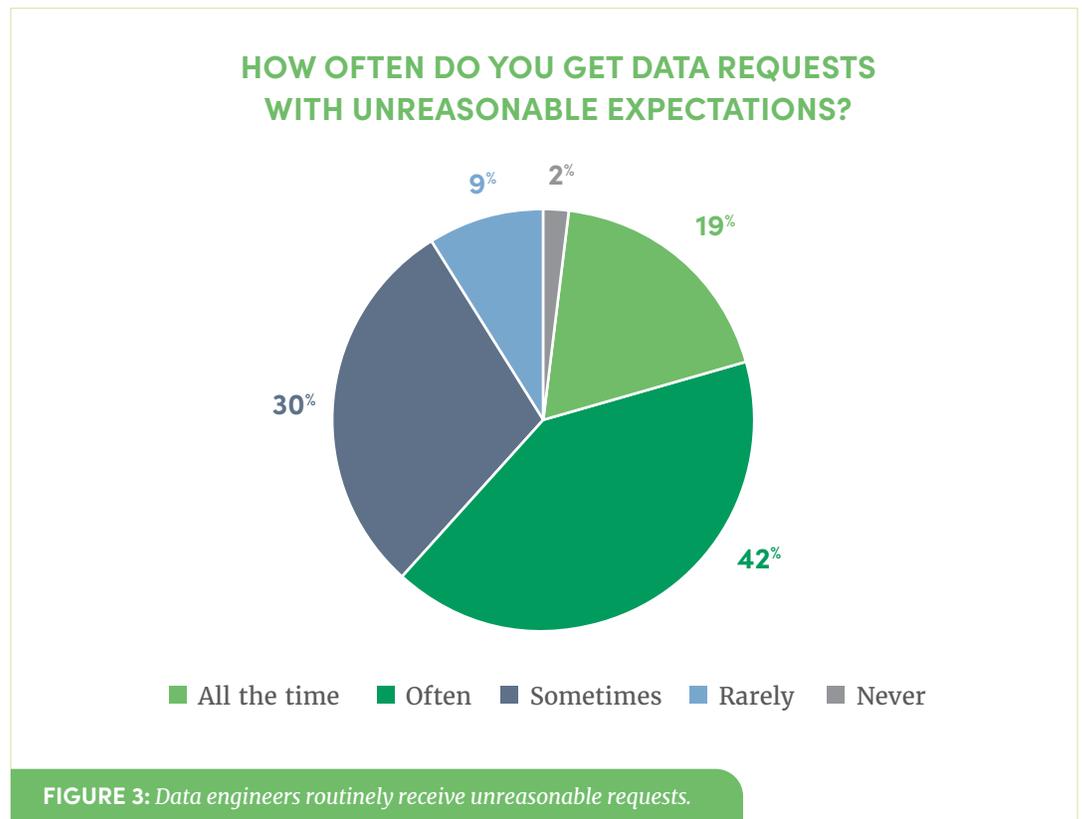
Companies that wish to cultivate data analytics as a sustainable competitive advantage need to find a way to flip the script. Data engineers need to be spending 80% of their time on value-add initiatives and 20% on operational execution and internal meetings.



YOU WANT IT WHEN?

Another major cause of burnout is the steady stream of half-baked requests from stakeholders. Like everything else in our *one-click* world, analytics is now expected to happen instantaneously and reliably. Ninety-one percent of our respondents reported receiving requests for analytics with unrealistic or unreasonable expectations. 61% said this happens “often” or “all the time” — figure 3. A majority of respondents reported receiving requests for analytics that are simply not possible to complete in the time requested or not possible with the functions and features specified. This phenomenon is not a surprise since business colleagues have little understanding of the complexity required to deliver accurate charts and graphs to decision-makers.

Data engineers focus on delivering clean and accurate data, and an effective way to optimize these impractical requests is to catalog enterprise data. By implementing a data catalog platform, data engineering teams can better understand and connect all data sources, simplifying managing and monitoring data pipelines. At the end of the day, unreasonable asks will happen, and a data catalog is the proper secret force to allow data engineers to meet crazy expectations.



SHAME AND BLAME

Other teams depend on the product of data engineering. Regular data outages erode trust, so when there's a problem in critical analytics, people within the organization engage in blaming and finger-pointing. In our survey, 87% of respondents said they are blamed frequently when things go wrong with the company's data and analytics. Sixty-three percent said this happens "often" or "all the time."

Public shaming can cause a range of bad feelings among data engineers. It can lead to anxiety and a reluctance to take technical risks — a significant obstacle to productivity. Shaming and blaming take the fun out of the most enjoyable aspects of data engineering — working with data.

STYMIED BY GOVERNANCE BUREAUCRACY

On the topic of working with data, 69% of those surveyed said their company's data governance policies make their day-to-day job more difficult. The "lock-it-down" approach employed by many organizations lacks transparency, often resulting in more work for data engineers who are beholden to complicated processes for managing access to data sources.

Enterprises can alleviate this burden by practicing [Agile Data Governance](#). Unlike traditional top-down data governance, Agile Data Governance opens up some traditionally restricted governance functions to a broader audience to iteratively capture the knowledge of data producers and consumers so everyone can benefit. Think of it like putting access on rails: making data fully auditable and predictable simplifies its management, so data engineers are free to work on more impactful projects.

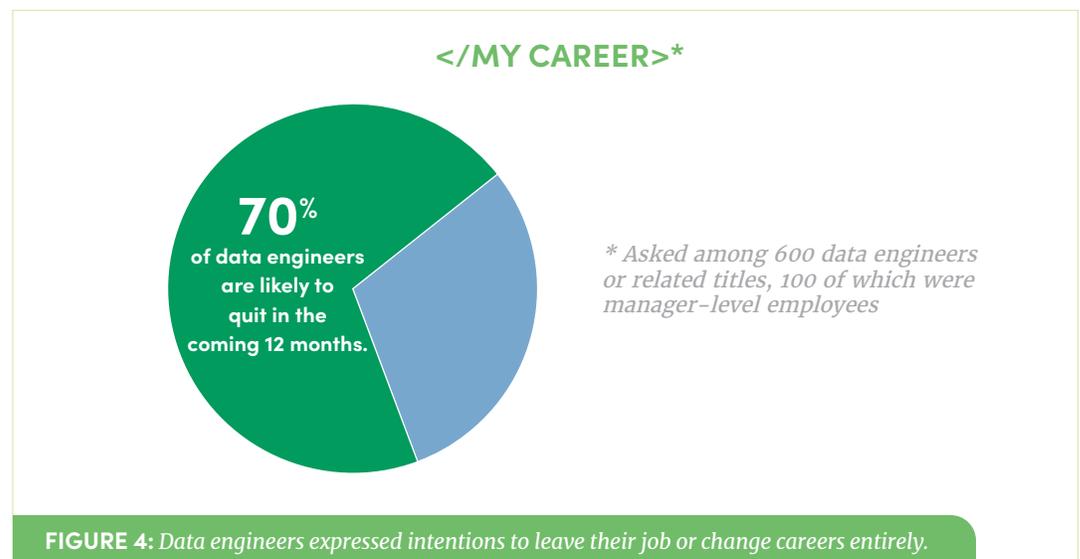
WORK-LIFE IMBALANCE

Data engineers work hard and deserve to spend their downtime relaxing. The problem is that issues and errors create unplanned work, which forces data engineers to work long, irregular schedules. Eighty-nine percent of data engineers in our survey reported frequent disruptions to work-life balance due to unplanned work. Fifty percent said this occurs "often" or "all the time."

Data engineers work overtime to compensate for the gap between performance and expectations. When a deliverable is met, data engineers are considered heroes. However, "heroism" is a trap. Heroes give up work-life balance. Yesterday's heroes are quickly forgotten when there is a new deliverable to meet. The long hours eventually lead to burnout, anxiety and even depression. Heroism is difficult to sustain over a long period, and it ultimately just resets expectations at a higher level without addressing the root cause of productivity bottlenecks.

I'M OUTTA HERE

In light of what we have learned about burnout, obstacles to productivity, high-profile shaming and lack of work-life balance, perhaps it isn't a surprise that over 70% of the data engineers surveyed indicated that they are likely to leave their current company in the next twelve months. Even more surprising is that 79% of those surveyed have considered abandoning the field of data engineering entirely — figure 4. This suggests that data engineers don't believe that the challenges that they experience are specific to their particular job situation or enterprise. The problems are industry-wide and can't be avoided simply by changing companies. These data engineers feel that the profession of data engineering is broken. Can it be fixed?



TOOLS SERVE WORKFLOWS

Tools vendors have learned that they can garner significant attention by claiming that their tool alone will solve a particular data problem. Still, our surveyed group of data engineers see through the hype. Over 89% agreed with the statement that “cutting edge tools for managing data and building analytics are ineffective without processes that deploy, monitor and manage analytics throughout the lifecycle.” In truth, tools are not an end in themselves. They serve lifecycle workflows.

According to quality pioneer W. Edwards Deming, 94% of problems are “[common cause variation](#).” To decrease this variation, you must focus on the system or process. This observation applies equally to factories that make widgets and data organizations that produce analytics. The best way to reduce waste and eliminate errors is to improve systemic processes and workflows. When quality methods, such as lean manufacturing, are applied to the end-to-end lifecycle of data and analytics, the term of art is called “[DataOps](#).” Overall, 78% feel DataOps is essential or very important to successfully manage data processes. This was even higher among data engineering managers, 91% of whom recognized the importance of DataOps — see Figure 5 below.

WHAT IS DATAOPS?

DataOps is a collection of technical practices, workflows, cultural norms, and architectural patterns that enable:

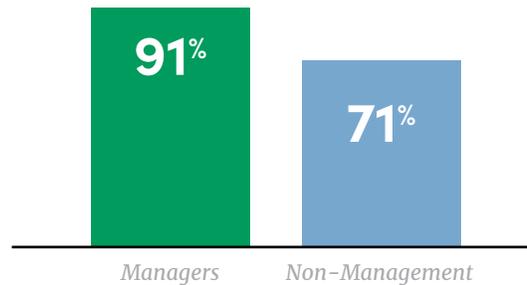
- Rapid innovation and experimentation, delivering new insights to customers with increasing velocity
- Extremely high quality and very low error rates
- Collaboration across complex arrays of people, technology, and environments
- End-to-end observability with clear and precise measurement, monitoring and transparency of results

In practical terms, DataOps employs automation to streamline analytics development and data operations workflows. DataOps also integrates testing, monitoring and observability into production data pipelines. By automating mundane tasks, DataOps enables data engineers to focus on higher-value activities. For example, data professionals can create sandbox environments for new analytics development projects on-demand — with a few clicks. DataOps slashes the time and effort required to release analytics to production by incorporating DevOps continuous integration, delivery and deployment into analytics development workflows. DataOps also institutes process and workflow metrics, so there is unprecedented transparency across the data lifecycle.

DataOps addresses each of the sources of burnout listed above. By introducing cataloging and observability into data operations and analytics development, DataOps helps the data team catch errors before they become critical outages. It automates data pipelines and prioritizes [last mile governance](#), democratizing data and improving collaboration between groups using hierarchies of orchestration. DataOps also slashes development cycle time by automating non-value-add aspects of analytics creation, enabling the data engineering and analytics team to keep up with the steady stream of requests by users. It also implements Agile Development, so the team stays focused on development that adds tangible value and receives immediate feedback from stakeholders.

TRUST THE PROCESSES

Overall, 78% feel DataOps is essential or very important to successfully manage data processes.



* Asked among 600 data engineers or related titles, 100 of which were manager-level employees

FIGURE 5: Data engineers recognize that DataOps is essential or very important to successfully manage data processes.

SAVING DATA ENGINEERS WITH DATAOPS AGILITY

The data engineers and managers in our survey spoke clearly about the challenges facing the data engineering profession. They feel burned-out by the relentless battle against data errors, inefficient manual processes, unreasonable requests, shaming and blaming, governance bureaucracy, and lack of work-life balance. The problem has reached a point where most individuals in the data engineering role are considering abandoning the profession altogether. Addressing data engineer burnout should be every organization's top priority.

Our survey participants also articulate a call to action. Enterprises can institute process change using methodologies like DataOps to orchestrate workflows that eliminate errors, accelerate delivery and deployment, foster collaboration and improve process transparency. By addressing the challenges of data engineering from a systemic and process perspective, DataOps can play a foundational role in enhancing the employee experience. Empowering people with processes and technology that enable greater communication, collaboration, integration, and automation will raise the quality and agility of analytics while putting the fun back into the data engineering role.

¹ Survey participant background: 600 data engineering respondents, 100+ of which were managers. 60.3% served in a data engineering role for 10+ years. Fifty-four percent reported 6+ years at their current company. Forty-seven percent work for companies in business for 26+ years. Fifty-three percent work for companies with 1000+ employees. Twenty-six percent of employers have less than \$50M in revenue. Seventeen percent of employers have more than \$500M in revenue. 84.8% come from industries outside software/IT.

² Erick Brethenoux, "Top Strategic Technology Trends for 2021: AI Engineering," Gartner ID: G00740659, December 30, 2020

³ Malody Chen, Nick Heudecker, "Survey Analysis: Data Management Struggles to Balance Innovation and Control," Gartner ID: G00464215,

**ABOUT
DATA.WORLD**

[data.world](#) is the enterprise data catalog for the modern data stack. Our cloud-native SaaS platform combines a consumer-grade user experience with a powerful knowledge graph to deliver enhanced data discovery, agile data governance, and actionable insights. data.world is a Certified B Corporation and public benefit corporation and home to the world's largest collaborative open data community with more than 1.3 million members. Our company has close to 50 patents and has been named one of Austin's Best Places to Work six years in a row. Follow us on [LinkedIn](#), [Twitter](#), and [Facebook](#), or [join us](#).

**ABOUT
DATAKITCHEN**

[DataKitchen](#) is the leader of the DataOps movement. It offers the only complete enterprise DataOps Platform that enables organizations to quickly, intuitively, and successfully implement and manage an end-to-end DataOps program using tools they already own. The Platform serves as the command center for DataOps. It simplifies complex toolchains, environments, and teams so that the entire data analytics organization can quickly innovate, seamlessly collaborate, and instantly deliver the kind of error-free, on-demand insight that leads to one successful business decision after another. To learn more about DataOps follow us on [LinkedIn](#) or [Twitter](#) or download the [DataOps Cookbook](#).